



Stein's Lemma—A Large Deviations Approach

R. SOWERS

*Target Characteristics Branch
Radar Division*

March 21, 1989

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			Approved for public release; distribution unlimited.		
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NRL Report 9185			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Naval Research Laboratory		6b. OFFICE SYMBOL (If applicable) Code 5340	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State, and ZIP Code) Washington, DC 20375-5000			7b. ADDRESS (City, State, and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 62712N	PROJECT NO.	TASK NO. XF12-141-100
					WORK UNIT ACCESSION NO. DN380-10
11. TITLE (Include Security Classification) Stein's Lemma—A Large Deviations Approach					
12. PERSONAL AUTHOR(S) Sowers, R.					
13a. TYPE OF REPORT Interim		13b. TIME COVERED FROM _____ TO _____		14. DATE OF REPORT (Year, Month, Day) 1989 March 21	
15. PAGE COUNT 14					
16. SUPPLEMENTARY NOTATION (See page ii)					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP			
			Stein's lemma Large deviations		
			Cramer's theorem		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>In this report, we prove Stein's Lemma by using a Large Deviations principle. Our proof is general, direct, and intuitive. We represent the log-likelihood ratio used to test between the two hypotheses on the basis of the first n observations as a sample mean of i.i.d. observations. Led by the Strong Law of Large Numbers, we formulate a series of hypothesis tests that bound the true Neyman-Pearson tests. We then determine the asymptotic behavior of these tests by using arguments from the proof of Cramer's Theorem. The conclusion of Stein's Lemma follows.</p>					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL R. Sowers			22b. TELEPHONE (Include Area Code) (202) 767-6321		22c. OFFICE SYMBOL Code 5340

16. SUPPLEMENTARY NOTATION

This work was performed while the author was with the Department of Electrical Engineering and Systems Research Center of the University of Maryland, College Park, MD 20742. The work of this author was supported through an ONR Graduate Fellowship. The author is currently with the Department of Mathematics, University of Maryland, College Park, MD 20742.

CONTENTS

INTRODUCTION	1
THE MAIN RESULT	1
MOTIVATION FOR THE PROOF OF STEIN'S LEMMA	2
PROOF OF THEOREM 1	3
CLOSURE	6
ACKNOWLEDGMENTS	7
REFERENCES	7
APPENDIX A — The Divergence Integral	8
APPENDIX B — The Strong Law of Large Numbers	9

STEIN'S LEMMA—A LARGE DEVIATIONS APPROACH

INTRODUCTION

In this report, we prove Stein's Lemma, see Ref. 1, by using a Large Deviations Principle. This idea was first proposed in Ref. 2; we provide a proof that is more general, direct, and intuitive.

Stein's Lemma is formulated as follows. Let $\{X_n\}_1^\infty$ be a sequence of i.i.d. observations defined on some underlying probability triple (Ω, \mathcal{F}, R) and taking values in a measurable space (E, \mathcal{E}) . We know that the probability measure R is one of two probability measures P or Q . For each $n = 1, 2, \dots$, we form a Neyman-Pearson test to decide whether $R = P$ or $R = Q$ on the basis of X_1, X_2, \dots, X_n (clearly, we need that $P \neq Q$ for this problem to be meaningful). Stein's Lemma states that for a fixed power constraint, the size of the Neyman-Pearson tests decays at an exponential rate and provides a formula for this rate.

To place the problem in a rigorous setting, let $\{\mathcal{F}_n\}_1^\infty$ be the filtration of \mathcal{F} generated by the observations;

$$\mathcal{F}_n := \sigma\{X_1, X_2 \dots X_n\}. \quad n = 1, 2, \dots \quad (1)$$

Let $0 < \epsilon < 1$ be a predetermined constant, and take $n = 1, 2, \dots$. For each set D in \mathcal{F}_n , we can define a decision rule to select P or Q by choosing P if and only if $\omega \in D$ for any $\omega \in \Omega$ (the requirement that D be in \mathcal{F}_n is of course equivalent to the requirement that our decision be a function of the observations $X_1, X_2 \dots X_n$). To form the Neyman-Pearson test of power ϵ , we vary $D \in \mathcal{F}_n$ so as to minimize the size $Q(D)$ (the false alarm rate in radar parlance) subject to the requirement that the power $P(D)$ satisfy $P(D) \geq 1 - \epsilon$ (i.e., a lower bound on the detection probability). Let $e(n, \epsilon)$ be this minimum, or more exactly, infimum; symbolically

$$e(n, \epsilon) := \inf\{Q(D) : D \in \mathcal{F}_n, P(D) \geq 1 - \epsilon\}. \quad n = 1, 2, \dots \quad (2)$$

Define \bar{P} (respectively \bar{Q}) as the probability measure induced on (E, \mathcal{E}) by any one of the observation RV's X_1, X_2, \dots under the probability measure P (respectively Q). Since the observations are identically distributed, it does not matter which X_n we select to define \bar{P} and \bar{Q} ; we may choose $\bar{P} = PX_1^{-1}$ and $\bar{Q} = QX_1^{-1}$. The result that we wish to prove can now be stated.

THE MAIN RESULT

Theorem 1 (Stein). Assume that \bar{P} is absolutely continuous with respect to \bar{Q} . Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log e(n, \epsilon) = -D(\bar{P}, \bar{Q}) \quad (3)$$

where

$$D(\bar{P}, \bar{Q}) := \int_E \log \frac{d\bar{P}}{d\bar{Q}} d\bar{P}, \quad (4)$$

the integral possibly being infinite.

If the observation space E is finite, this result is the same as the one in Ref. 3, Corollary 2.2.2, and in Ref. 2. We note that $D(\bar{P}, \bar{Q})$ is the Kullback-Leibler informational divergence of \bar{P} from \bar{Q} ; thus we know

that $D(\bar{P}, \bar{Q})$ is well defined; see Ref. 4 and Appendix A. Note also that if \bar{P} is not absolutely continuous with respect to \bar{Q} , the Neyman-Pearson tests are trivial and $e(n, \epsilon) = 0$ for n large. Indeed, assume that \bar{P} is not absolutely continuous with respect to \bar{Q} so that there is a set A in \mathcal{E} such that $\bar{P}(A) > 0$ but $\bar{Q}(A) = 0$. For each $n = 1, 2, \dots$, define the decision region $D_n \in \mathcal{F}_n$ as

$$\begin{aligned} D_n &:= \{X_i \in A \text{ for some } i = 1, 2, \dots, n\} \\ &= \bigcup_{i=1}^n \{X_i \in A\}. \end{aligned} \quad (5)$$

Then

$$\begin{aligned} P(D_n) &= 1 - P(\bar{D}_n) \\ &= 1 - P\left(\bigcap_{i=1}^n \{X_i \in \bar{A}\}\right) \\ &= 1 - \bar{P}(\bar{A})^n, \end{aligned} \quad (6)$$

so that $\lim_n P(D_n) = 1$, and consequently D_n satisfies the power constraint for n large. But since

$$\begin{aligned} Q(D_n) &\leq \sum_{i=1}^n Q\{X_i \in A\} \\ &= \sum_{i=1}^n \bar{Q}(A) = 0, \end{aligned} \quad (7)$$

we conclude that for n large, $e(n, \epsilon) \leq Q(D_n) = 0$.

MOTIVATION FOR THE PROOF OF STEIN'S LEMMA

It is a well-known result that Neyman-Pearson tests are performed by comparing a log-likelihood ratio to a threshold, see Ref. 5, Thm. 5.5.2. For each $n = 1, 2, \dots$, let P_n (respectively Q_n) be the restriction of P (respectively Q) to the σ -field \mathcal{F}_n . The absolute continuity requirement on \bar{P} and \bar{Q} implies that for each $n = 1, 2, \dots$, P_n is absolutely continuous with respect to Q_n , so that our log-likelihood ratio is $\log dP_n/dQ_n$. If we define

$$Y_i := \log \frac{d\bar{P}}{d\bar{Q}}(X_i), \quad n = 1, 2, \dots \quad (8)$$

then it is not difficult to verify that

$$\log \frac{dP_n}{dQ_n} = S_n \quad n = 1, 2, \dots \quad (9)$$

where

$$S_n := \sum_{i=1}^n Y_i. \quad n = 1, 2, \dots \quad (10)$$

Note that the sequence $\{Y_n\}_1^\infty$ is an i.i.d. sequence and that we have suggestively written the log-likelihood ratio as a partial sum. If $R = P$, then by the Strong Law of Large Numbers (SLLN),

$$\begin{aligned} \frac{1}{n} S_n &\xrightarrow{P\text{-a.s.}} \int_{\Omega} Y_1 dP \\ &= \int_E \log \frac{d\bar{P}}{d\bar{Q}} d\bar{P} \\ &= D(\bar{P}, \bar{Q}). \end{aligned} \quad (11)$$

Alternately, if $R = Q$, we would then expect that if \bar{Q} is absolutely continuous with respect to \bar{P} ,

$$\begin{aligned} \frac{1}{n} S_n &\xrightarrow{Q\text{-a.s.}} \int_{\Omega} Y_1 dQ \\ &= \int_E \log \frac{d\bar{P}}{d\bar{Q}} d\bar{Q} \\ &= - \int_E \log \frac{d\bar{Q}}{d\bar{P}} d\bar{Q} \\ &= -D(\bar{Q}, \bar{P}). \end{aligned} \tag{12}$$

(See Appendix B for the SLLNs that are required in Eqs. (11) and (12) if the integrals are infinite.)

From Eqs. (11) and (12), our hypothesis tests should reflect the fact that S_n/n has different almost sure limits under the different probability measures P and Q . If we define our decision regions so as to decide that $R = P$ if S_n/n is near $D(\bar{P}, \bar{Q})$, then a high rate of detection and a low false alarm rate should result for large n . (Note that Eqs. (11) and (12) explain a technical difficulty. If P agrees with Q on $\mathcal{F}_{\infty} := \bigvee_{n=1}^{\infty} \mathcal{F}_n$, then we expect not to be able to distinguish between $R = P$ and $R = Q$ from the observations. This is reflected in the easily verified fact that $\bar{P} = \bar{Q}$ if and only if P and Q coincide on \mathcal{F}_{∞} , in which case $D(\bar{P}, \bar{Q}) = D(\bar{Q}, \bar{P}) = 0$ and S_n/n tends almost surely to 0 under both P and Q .) Since $\{Y_i\}_1^{\infty}$ is i.i.d., we can use Cramér's theorem from the field of Large Deviations, see Ref. 6, Theorem 3.8 and Ref. 7, Theorem 3.1, to describe the *rate* at which S_n/n tends to its limit under P and Q . The reasoning behind the following arguments is then clear.

PROOF OF THEOREM 1

A Large Deviations Principle

Let us temporarily assume that

- (a) the probability measure \bar{Q} is absolutely continuous with respect to the probability measure \bar{P} ,
- (b) $D(\bar{P}, \bar{Q}) < \infty$ and $D(\bar{Q}, \bar{P}) < \infty$, and
- (c) the moment generating function M of Y_1 under Q ; i.e., $M(\theta) := \int_{\Omega} e^{\theta Y_1} dQ$, is finite for all θ in \mathbb{R} .

Under these assumptions, we may directly verify the upper bound

$$\limsup_n \frac{1}{n} \log e(n, \epsilon) \leq -D(\bar{P}, \bar{Q}) \tag{13}$$

by invoking Cramér's Theorem. Fix $\delta > 0$ and set $F_{\delta} := [D(\bar{P}, \bar{Q}) - \delta, \infty)$. From Eq. (11) we know that if $R = P$, then $\lim_n S_n/n \in F_{\delta}$ P -a.s. Since almost sure convergence is stronger than convergence in probability, it is immediate that

$$\lim_n P(\{S_n/n \in F_{\delta}\}) = 1, \tag{14}$$

so for large n , the decision regions given by

$$D_n := \{S_n/n \in F_{\delta}\} \quad n = 1, 2, \dots \tag{15}$$

satisfy the power constraint.

We can now apply Cramér's Theorem to verify that

$$\limsup_n \frac{1}{n} \log Q(D_n) \leq - \inf_{x \in F_{\delta}} I(x), \tag{16}$$

where I is the Legendre-Fenchel transform, see Ref. 8, Chapter 6, of $\log M(\cdot)$;

$$I(x) := \sup_{\theta \in \mathbb{R}} (\theta x - \log M(\theta)). \quad x \in \mathbb{R} \quad (17)$$

From Ref. 6, Lemma 3.3, we know that $x \mapsto I(x)$ is nondecreasing for $x \geq \int_{\Omega} Y_1 dQ = -D(\bar{Q}, \bar{P})$, the integral being well defined under assumption (b). Thus

$$\inf_{x \in F_{\delta}} I(x) = I(D(\bar{P}, \bar{Q}) - \delta). \quad (18)$$

But

$$\begin{aligned} I(D(\bar{P}, \bar{Q}) - \delta) &\geq (1)(D(\bar{P}, \bar{Q}) - \delta) - \log M(1) \\ &= D(\bar{P}, \bar{Q}) - \delta \end{aligned} \quad (19)$$

since

$$M(1) = \int_E e^{\log \frac{dP}{dQ}} d\bar{Q} = \bar{P}(E) = 1. \quad (20)$$

Combining Eqs. (16) through (19), we have that

$$\limsup_n Q(D_n) \leq -D(\bar{P}, \bar{Q}) + \delta. \quad (21)$$

In view of Eq. (14), we then have that

$$\limsup_n \frac{1}{n} \log e(n, \epsilon) \leq \limsup_n \frac{1}{n} \log Q(D_n) \leq -D(\bar{P}, \bar{Q}) + \delta. \quad (22)$$

Since $\delta > 0$ was arbitrary, Eq. (13) is established.

An inspection of the proof of Cramér's theorem reveals how to prove Theorem 1 when assumptions (a) through (c) are not enforced.

Case 1: $D(\bar{P}, \bar{Q}) < \infty$

Upper Bound: Fix $\delta > 0$ and again set $F_{\delta} := [D(\bar{P}, \bar{Q}) - \delta, \infty)$ and

$$D_n := \{S_n/n \in F_{\delta}\}. \quad n = 1, 2, \dots \quad (23)$$

As in the above arguments, we know that for large n , D_n satisfies the power constraint. Following the arguments of Ref. 6, Lemma 3.4, we argue that for each $n = 1, 2, \dots$

$$\begin{aligned} Q(D_n) &= \int_{\{S_n \geq n(D(\bar{P}, \bar{Q}) - \delta)\}} dQ \\ &\leq \int_{\Omega} \exp [S_n - n(D(\bar{P}, \bar{Q}) - \delta)] d\bar{Q} \\ &= \exp [-n(D(\bar{P}, \bar{Q}) - \delta)] \int_{\Omega} e^{S_n} dQ \\ &= \exp [-n(D(\bar{P}, \bar{Q}) - \delta)] (1), \end{aligned} \quad (24)$$

and consequently,

$$\limsup_n \frac{1}{n} \log Q(D_n) \leq -D(\bar{P}, \bar{Q}) + \delta. \quad (25)$$

As above, this is sufficient to prove the upper bound Eq. (13) since $\delta > 0$ was arbitrary.

Lower Bound: We next prove that

$$\liminf_n \frac{1}{n} \log e(n, \epsilon) \geq -D(\bar{P}, \bar{Q}). \quad (26)$$

Our proof of Eq. (26) is essentially the same as that in Ref. 2. Take $\delta > 0$. Then for each positive integer n , we can find a set U_n in \mathcal{F}_n so that

$$P(U_n) \geq 1 - \epsilon \quad (27)$$

and

$$Q(U_n) \leq e(n, \epsilon) e^{n\delta/2}. \quad (28)$$

Define $F_\delta := (-\infty, D(\bar{P}, \bar{Q}) + \delta/2]$ and set

$$D_n := \{S_n/n \in F_\delta\}. \quad n = 1, 2, \dots \quad (29)$$

As in the proof of the upper bound, the SLLN ensures that $\lim_n P(D_n) = 1$, so necessarily

$$\liminf_n P(U_n \cap D_n) \geq 1 - \epsilon. \quad (30)$$

For each $n = 1, 2, \dots$,

$$\begin{aligned} P(U_n \cap D_n) &= P_n(U_n \cap D_n) \\ &= \int_{U_n \cap D_n} e^{S_n} dQ \\ &\leq \int_{U_n \cap D_n} \exp [n (D(\bar{P}, \bar{Q}) + \delta/2)] dQ \\ &\leq \exp [n (D(\bar{P}, \bar{Q}) + \delta/2)] Q(U_n \cap D_n), \end{aligned} \quad (31)$$

where we have used Eq. (9) and the fact that $S_n \leq n (D(\bar{P}, \bar{Q}) + \delta/2)$ on D_n , which is obvious from Eq. (29). Thus, upon combining Eq. (28) and Eq. (31), we have

$$\begin{aligned} e(n, \epsilon) &\geq Q(U_n) e^{-n\delta/2} \\ &\geq Q(U_n \cap D_n) e^{-n\delta/2} \\ &\geq P(U_n \cap D_n) \exp [-n (D(\bar{P}, \bar{Q}) + \delta)], \end{aligned} \quad (32)$$

so in view of Eq. (30),

$$\liminf_n \frac{1}{n} \log e(n, \epsilon) \geq -D(\bar{P}, \bar{Q}) - \delta; \quad (33)$$

since $\delta > 0$ was arbitrary, Eq. (26) is true.

Case 2: $D(\bar{P}, \bar{Q}) = \infty$

We wish to prove that

$$\liminf_n \frac{1}{n} \log e(n, \epsilon) = -\infty. \quad (34)$$

From the SLLN found in Appendix B, we know that $\lim_n S_n/n = \infty$ P -a.s. if $R = P$. Fix a positive number B , and define $F_B := [B, \infty)$ and for each $n = 1, 2, \dots$, let the decision region D_n be given by

$$D_n := \{S_n/n \in F_B\}. \quad (35)$$

Then $\lim_n P(D_n) = 1$ as in Case 1, but for each $n = 1, 2, \dots$,

$$\begin{aligned} Q(D_n) &= \int_{\{S_n \geq nB\}} dQ \\ &\leq \int_{\Omega} \exp[S_n - nB] dQ \\ &= e^{-nB} \int_{\Omega} e^{S_n} dQ \\ &= e^{-nB}(1), \end{aligned} \tag{36}$$

so that

$$\limsup_n \frac{1}{n} \log Q(D_n) \leq -B. \tag{37}$$

Hence, in a manner analogous to Eq. (22),

$$\limsup_n \frac{1}{n} \log e(n, \epsilon) \leq \limsup_n \frac{1}{n} \log Q(D_n) \leq -B, \tag{38}$$

and since B was an arbitrary positive number,

$$\limsup_n \frac{1}{n} \log e(n, \epsilon) = -\infty, \tag{39}$$

which was to be proved.

The proof of Theorem 1 is complete.

CLOSURE

Note in our proof of Stein's Lemma that we did not formulate the Neyman-Pearson tests. The Strong Law of Large Numbers and Eqs. (11) and (12) led us to a series of tests that bounded the true Neyman-Pearson tests. The asymptotic behavior of these tests was found by using Large Deviations arguments, and Stein's Lemma resulted.

In the Neyman-Pearson tests studied here, we minimized the false alarm rate subject to a lower bound on the probability of detection. The more common formulation is to maximize the probability of detection subject to an upper bound on the false alarm rate. By reversing the roles of P and Q , we see that the two problems are equivalent. Define

$$\gamma(n, \epsilon) := \sup\{Q(D) : D \in \mathcal{F}_n, P(D) \leq \epsilon\}. \tag{40}$$

Then $\gamma(n, \epsilon)$ corresponds to maximizing the probability of detection $Q(D)$ (the power) subject to the constraint that the false alarm rate $P(D)$ satisfy $P(D) \leq \epsilon$ (an upper bound on the size). Since

$$\gamma(n, \epsilon) = 1 - \epsilon(n, \epsilon), \tag{41}$$

an alternate way of stating the result of Theorem 1 is

$$\lim_n \frac{1}{n} \log (1 - \gamma(n, \epsilon)) = -D(\bar{P}, \bar{Q}). \tag{42}$$

Note also that we can in fact relax the assumption that the observations are i.i.d. If P_n is absolutely continuous with respect to Q_n for each $n = 1, 2, \dots$, and if there is a constant M , possibly infinite, such that

$$M = P\text{-}\lim_n \frac{1}{n} \log \frac{dP_n}{dQ_n}, \quad (43)$$

then it is easy to verify, using the above arguments, that

$$\lim_n \frac{1}{n} \log e(n, \epsilon) = -M. \quad (45)$$

We shall leave the proof of this extension to the interested reader.

ACKNOWLEDGMENTS

This work was made possible by an Office of Naval Research Graduate Fellowship. Part of the work was performed at the Naval Research Laboratory, Code 5340. The author's ONR Fellowship mentor, while at NRL, was Dr. Frank Kretschmer, Jr. The author is currently with the Department of Mathematics and Systems Research Center of the University of Maryland; he thanks Dr. Pantelis Tsoucas of the University of Maryland for introducing him to Anantharam's paper, and Dr. Armand Makowski of the University of Maryland, Dr. John Gubner of the University of Wisconsin, and James Sangston of the Naval Research Laboratory for their helpful suggestions.

REFERENCES

1. H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on a Sum of Observations," *Ann. Math. Stat.* **23**, 493–507 (1952).
2. V. Anantharam, "A Large Deviations Approach to Error Exponents in Source Coding and Hypothesis Testing," Preprint, Cornell University, Ithaca, NY.
3. I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Inc., New York, 1981.
4. S. Kullback and R. Leibler, "On Information and Sufficiency," *Ann. Math. Stat.* **22**, 79–86 (1951).
5. J. Barra and L. L. Herbach, *Mathematical Basis of Statistics*, Academic Press, Inc., New York, 1981.
6. D. Stroock, *An Introduction to the Theory of Large Deviations*, Springer-Verlag, New York, 1984.
7. S. Varadhan, *Large Deviations and Applications*, SIAM, Philadelphia, 1984.
8. R. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Springer-Verlag, New York, 1985.
9. R. Laha and V. Rohatgi, *Probability Theory*, John Wiley & Sons, 1979.

Appendix A

THE DIVERGENCE INTEGRAL

The proof that the divergence integral is well defined is difficult to find in the literature; here we provide a simple proof. We show that

$$\int_E \left(\log \frac{d\bar{P}}{d\bar{Q}} \right)^- d\bar{P} < \infty, \quad (\text{A1})$$

where $x^- := \min\{-x, 0\}$. For convenience, define $X := \frac{d\bar{P}}{d\bar{Q}}$. Now

$$\begin{aligned} \int_E (\log X)^- d\bar{P} &= \int_{\{X \leq 1\}} -\log X d\bar{P} \\ &= \int_{\{X \leq 1\}} -X \log X d\bar{Q} \\ &= \int_{\{X \leq 1\}} \varphi(X) d\bar{Q}, \end{aligned} \quad (\text{A2})$$

where $\varphi(t) := -t \log t$ for $t > 0$ and $\varphi(0) = 0$. If $\bar{Q}\{X \leq 1\} = 0$, Eq. (A1) follows immediately from Eq. (A2), so assume that $\bar{Q}\{X \leq 1\} > 0$. By differentiating twice, we see that φ is concave on $[0, \infty)$, so by Jensen's inequality,

$$\begin{aligned} \int_{\{X \leq 1\}} \varphi(X) d\bar{Q} &\leq \bar{Q}\{X \leq 1\} \varphi \left(\frac{1}{\bar{Q}\{X \leq 1\}} \int_{\{X \leq 1\}} X d\bar{Q} \right) \\ &= \bar{Q}\{X \leq 1\} \varphi \left(\frac{\bar{P}\{X \leq 1\}}{\bar{Q}\{X \leq 1\}} \right), \end{aligned} \quad (\text{A3})$$

which is clearly finite; returning to Eq. (A2), we see that Eq. (A1) is true when $\bar{Q}\{X \leq 1\} > 0$.

Appendix B

THE STRONG LAW OF LARGE NUMBERS

The proof of Eqs. (11) and (12) requires the following formulation of the Strong Law of Large Numbers.

Proposition. Let $\{X_n\}_1^\infty$ be a sequence of i.i.d. RVs defined on an underlying probability triple (Ω, \mathcal{F}, P) . Suppose that $E[X_1]$ is well defined and $-\infty < E[X_1] \leq \infty$. Then

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P\text{-a.s.}} E[X_1]. \quad (B1)$$

Proof. If $E[X_1] < \infty$, we may use Ref. 9, Theorem 2.3.1 to verify Eq. (B1); assume that $E[X_1] = \infty$. Take any positive constant B , and define

$$X_n^B := \min\{X_n, B\}. \quad n = 1, 2, \dots \quad (B2)$$

Clearly the $\{X_n^B\}_1^\infty$ are i.i.d. and P -integrable, so Ref. 9, Theorem 2.3.1 again applies and we conclude that P -a.s.

$$\lim_n \frac{1}{n} \sum_{i=1}^n X_i^B = E[X_1^B]. \quad (B3)$$

But P -a.s.

$$\liminf_n \frac{1}{n} \sum_{i=1}^n X_i \geq \liminf_n \frac{1}{n} \sum_{i=1}^n X_i^B \geq E[X_1^B]. \quad (B4)$$

Since B was an arbitrary positive constant, we let B tend to infinity, and by the Monotone Convergence Theorem, we then have from Eq. (B4) that P -a.s.

$$\liminf_n \frac{1}{n} \sum_{i=1}^n X_i = \infty, \quad (B5)$$

which is the result we seek when $E[X_1] = \infty$. ■

